



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Reduplicated words are easier to learn

Citation for published version:

Ota, M & Skarabela, B 2016, 'Reduplicated words are easier to learn', *Language Learning and Development*, vol. 12, no. 4, pp. 380-397. <https://doi.org/10.1080/15475441.2016.1165100>

Digital Object Identifier (DOI):

[10.1080/15475441.2016.1165100](https://doi.org/10.1080/15475441.2016.1165100)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Language Learning and Development

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Reduplicated Words Are Easier to Learn

Mitsuhiko Ota and Barbora Skarabela

University of Edinburgh

Author Note

Correspondence concerning this article should be addressed to Mitsuhiko Ota, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, 3 Charles Street, Edinburgh EH8 9AD, UK.

Acknowledgment: The authors wish to thank Judit Fazekas for help with data collection and coding, and the parents and infants who participated in this study.

Funding: This work was funded by an Economic and Social Research Council grant (ES/J023825/1) awarded to Mitsuhiko Ota.

Email: mits@ling.ed.ac.uk

Abstract

Infants' disposition to learn repetitions in the input structure has been demonstrated in pattern generalization (e.g., learning the pattern ABB from the token *ledidi*). This study tested whether a repetition advantage can also be found in lexical learning (i.e., learning the word *lele* vs. *ledi*). Twenty-four English-learning infants (mean age: 18.5 months) were exposed to novel word-object mappings involving either a reduplicated CVCV word (e.g., *neenee*) or a nonreduplicated CVCV word (e.g., *bolay*). Infants were more adept at learning word-object mappings with a reduplicated word than with a nonreduplicated word. A follow-up corpus analysis of infant-directed speech showed that this preference could not be attributed to the frequency patterns of reduplicated words or syllables in the linguistic input. These findings indicate that an experience-independent bias toward repeated elements plays an important role not only in pattern generalization but also in word learning.

Keywords: word learning, learning bias, reduplication, syllable repetition, infant-directed speech

Reduplicated Words Are Easier to Learn

One of the central questions in language acquisition research is whether infants have predispositions to preferentially attend to and process certain aspects of the input from the linguistic environment. Some such predispositions, or biases, may be specific to the domain of language, making the infant learner attend to particular linguistic configurations in the input that are determined a priori by universal restrictions on language structures (Chomsky, 1965; Wexler & Culicover, 1980). Other biases may come from limits in human perception (Creel, Newport, & Aslin, 2004), attention (Toro, Sinnett, & Soto-Faraco, 2005; Pacton & Perruchet, 2008), or working memory (Newport, 1988) that apply to domain-general learning mechanisms employed in the context of language acquisition (e.g., statistical learning). In recent years, there has also been growing interest in the possibility that humans are equipped with cognitive or perceptual biases that are not unique to language processing but are nonetheless attuned to specific aspects of linguistic input or particular types of regularities that underlie the surface input (de Vries et al., 2012; Endress, Nespor, & Mehler, 2009; Feldman, 2003; Moreton, 2012; Saffran, 2002; Saffran & Thiessen, 2003). For example, it has been proposed that, across domains, humans are more likely to learn patterns involving repetition (Endress, Dahan-Lambertz, & Mehler, 2007), edges of strings (Endress, Scholl, & Mehler, 2005), intra-dimensional dependencies (Moreton, 2012), predictive dependencies (Saffran, 2002), and minimal Boolean complexity (Feldman, 2000). These types of learning biases may not have evolved in humans to serve language acquisition *per se*, but they may still be critically involved in the process of language development by virtue of the constraints they impose on the potentially vast range of computations that a general learning mechanism can perform on the available linguistic information.

The current study focuses on one such bias, or learning primitive, proposed in the literature: the tendency to attend to identity relations or repetitions in the input, particularly

adjacent repetitions (Endress et al., 2007, 2009; Gervain & Werker, 2008). This proposal is consistent with a number of related findings. Neonates display stronger neural response to trisyllabic sequences with adjacent repetition (e.g., *mubaba*) compared to no, or non-adjacent, repetition (e.g., *mubage*, *bamuba*) (Gervain et al., 2008, 2012). Repetition-based generalizations are readily learned by infants and adults, whether the stimuli are linguistic (e.g., syllables), visuo-linguistic (e.g., letters), or nonlinguistic (e.g., musical notes) (Marcus et al., 1999; Endress et al., 2007; Gerken, 2010, Gomez & Gerken, 1999). Furthermore, learners' performance in artificial grammar learning tasks deteriorates noticeably when repetition patterns are removed from the strings (Gomez, Gerken, & Schvaneveldt, 2000; Tunney & Altmann, 2001).

Much previous research on the role of repeated elements in language acquisition has been carried out as part of an investigation into the detection and abstraction of underlying structures in the input. For example, infants in Marcus et al. (1999) were exposed to exemplars such as *ga-ti-ti* and *li-na-na*, and tested on their ability to generalize the ABB pattern to other tokens with the same structure (e.g., *wo-fe-fe*). This process of learning can be understood as extraction of abstract patterns involving repetition, or *type repetitions* (Endress et al., 2005). However, there is another kind of learning process involving repetition, *token repetitions* (Endress et al., 2005), wherein the learning concerns detecting, extracting and/or storing repetition of the actual tokens in the input. To use the same examples from above, this entails learning the exact sequence of syllables containing repetition (e.g., *ga-ti-ti*), as opposed to either learning a sequence of syllables without repetition (e.g., *ga-ti-fe*) or generalizing the underlying pattern of repetition to other tokens (e.g., *wo-fe-fe*). Compared to type repetition, the role of token repetition in early language learning is much less understood.

One area where token repetition may be privileged is lexical learning. There are readily recognizable indications that adjacent token repetition plays some potential role in the composition of early vocabulary that infants learn or receive (Gervain & Werker, 2008; Endress et al., 2009). For example, reduplicated lexical items are often among children's first words (e.g., *bébé* 'baby' in French, *papa* 'daddy' in Italian, *tata* 'granddad' in Hungarian, and *baa-baa*, *yum-yum*, and *night-night* in English) (Gervain & Werker, 2008; see also Caselli et al., 1995; Tardif et al., 2008). Furthermore, across languages, adjacent repetition of a syllable (e.g., *choochoo*, *night-night*) is one of the most commonly attested features of 'baby-talk words', or the register-specific lexical items in infant-directed speech (Ferguson, 1964, 1978). While it is not clear to what extent the presence of such words changes the overall phonological profile of the lexical input to the young learner, the crosslinguistic tendency for register-specific items to contain repeated syllables may reflect a learning primitive that favors token repetition.

Somewhat surprisingly, few studies to this date have systematically examined whether words containing repetition are more readily learned. The most relevant experimental evidence known to us is an incidental finding reported by Bird, Chapman, and Schwartz (2004). The main purpose of Bird et al.'s (2004) study was to compare fast-mapping abilities of adolescents with Down Syndrome to those of mental-age matched controls, who were 4 to 6 years old. In a story retelling task, both groups reproduced more novel reduplicated CVCV words (e.g., *wayway*, *soosoo*) than CVC words (e.g., *tash*, *poom*), suggesting a learnability advantage of reduplicated words. However, findings from 4- to 6-year-olds may not apply to lexical learning in infants in the early stages of word learning. In addition, the results from this production task leave ambiguity as to whether the children actually learned the reduplicated CVCV words better than the CVC words, or whether they were more willing to use them in the retelling because of their ease of production. Moreover,

a comparison between reduplicated CVCV words and CVC words would not allow us to disentangle the role of reduplication from that of CV structure and the number of syllables in novel word learning. Thus, these findings do not constitute firm evidence that there is a mechanism that preferentially learns words with adjacent repeated syllables.

The main goal of this study, therefore, was to test the prediction that adjacent token repetition is privileged in the context of lexical acquisition, by contrasting infants' learning of reduplicated and nonreduplicated words. Specifically, English-learning infants around 18 months of age were exposed to two types of novel words: one with the structure $C_1V_1C_1V_1$ (e.g., *neenee* /nini/, *laylay* /lele/) and the other, with the structure $C_1V_1C_2V_2$ (e.g., *foonee* /funi/, *bolay* /bole/). The novel words were paired with unfamiliar objects. The prediction was that infants should be better at learning object-word mappings involving reduplicated words (i.e., $C_1V_1C_1V_1$) than nonreduplicated words (i.e., $C_1V_1C_2V_2$).

We note that demonstrating the advantage of reduplicated words in early lexical learning does not necessarily constitute evidence for a token repetition bias that is inherent and independent of experience. It is possible that infants' preference for learning words with repeated syllables may simply be a product of patterns in the linguistic environment, particularly given the observation mentioned above that reduplication is a common characteristic of lexical items unique to infant-directed speech. There is compelling evidence that by 9 months, infants become sensitive to the regular or predominant prosodic and phonotactic patterns of words in the ambient language, and begin to show a tendency to detect and learn words that match those phonological structures (Curtin, Mintz, & Christiansen, 2005; Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Houston, & Newsome, 1999; Storkel, 2009). Some phonological biases in lexical learning proposed earlier in the literature (e.g., a bias toward labial-coronal sequences) have in fact turned out to be language-specific effects that reflect the overall phonological or lexical properties of the ambient language

(Gonzalez-Gomez et al., 2014; Højen & Nazzi, 2016). Similarly, if in fact the typical linguistic input to the learner contains a disproportionate number of reduplicated words or adjacent sequences of identical syllables, then exposure to such input may result in an acquired bias toward learning reduplicated words more readily.

In order to investigate whether infants' preferential learning of reduplicated words could be attributed to an inherent, rather than acquired bias, a corpus analysis of infant-directed speech was carried out, examining whether English-learning infants are exposed to linguistic input containing a large proportion of reduplicated words or repetitions of syllables beyond the level expected by chance. The outcomes of this analysis would help to determine whether the results of the experiment could be understood as an effect of an experience-independent bias for repetition or a manifestation of the infants' prior exposure to a distribution pattern in the input language that may be skewed toward reduplicated words or syllable repetitions.

Experiment

Infants around 18 months of age are capable of learning two novel words or labels for separate unfamiliar referents after a short laboratory exposure (Houston-Price, Plunkett, & Harris, 2005; Nazzi, 2005; Schafer & Plunkett, 1998; Werker et al., 1998). Houston-Price et al. (2005) demonstrated this capacity in 18-month-olds who were familiarized to two novel nonreduplicated CVCV words (*shoofy* [ʃufi] and *gopper* [gɒpə], with the latter pronounced with no postvocalic /r/), and were then tested in a standard preferential looking paradigm. In our experiment, we trained infants of the same age on two different types of novel CVCV words, one reduplicated (e.g., *neenee*), and the other, not reduplicated (e.g., *foonee*). Learning of the two types of words was then measured separately through a modified preferential looking paradigm (Halberda, 2003; Killing & Bishop, 2008). In this paradigm, test trials began by displaying two objects (the target and a competitor) as the test word is played in a

carrier sentence “*Where is the ____?*”. This phase of the test trial was followed by an ‘answer’ phase during which only the target object was shown moving rapidly accompanied by the sentence “*There’s the ____!*” The prediction was that infants would look more to the target image of the trained reduplicated novel word compared to the trained nonreduplicated novel word before the ‘answer’ phase. As a basis of comparison, we also included real words familiar to the participating infants (e.g., *doggy*, *apple*). The target response pattern to these familiar words in the test phase would provide us with a baseline for infants’ behavior in this paradigm.

Method

Participants. Data were collected from 24 healthy infants (14 males, 10 females) approximately 18 months of age. They ranged in age from 17 month 2 days to 19 months 23 days, with a mean of 18 months 15 days. An additional 12 infants were tested, but excluded from the analysis for the following reasons: fussiness (N=6), equipment failure (N=1), failure to complete the study (N=1), and not reaching a 60% fixation criterion on at least one trial per condition (details of the criterion are given in the Results section) (N=4). The infants were recruited via local parent-toddler groups, magazine/leaflet advertisement, mailing lists, and word of mouth. All participants had normal hearing and vision, and were raised in an English-only or English dominant language environment in Scotland, UK. The infants were accompanied by a caregiver who received a £5 coffee voucher for their participation.

Materials. All auditory stimuli were read by a female native speaker of Standard Scottish English in a child-directed speech register, and recorded in a sound attenuated studio at a sampling rate of 44 kHz. The talker was blind to the research questions. Speech stimuli used in the familiarization phase consisted of the following carrier phrases, played with one of the experimental words:

1. *Look! A _____. Yes, that’s a _____!*

2. *Look! A _____. See? A _____!*

3. *Oh! A _____. Yes, that's a _____!*

Speech stimuli used in the test phase consisted of the following carrier phrases, played with one of the test items:

1. *Where's the _____? That's the _____!*

2. *Where's the _____? There's the _____!*

The experimental words embedded in these phrases comprised four real words and eight novel words. The real words (*doggy*, *bunny*, *balloon*, and *apple*) were known to have a comprehension rate of 75% or more at 18 months in a norming study (Dale & Fenson, 1996). Further confirmation was received from the parents prior to the experiment that their children understood these words. The novel words included four reduplicated CVCV forms (*neenee* /nini/, *foofoo* /fufu/, *bobo* /bobo/, and *laylay* /lele/) and four nonreduplicated CVCV forms consisting of the same syllables in the reduplicated forms (*neefoo* /nifu/, *foonee* /funi/, *bolay* /bole/, and *laybo* /lebo/). All novel words and familiar words had stress on the initial syllable. Comparisons of the phonotactic, neighborhood and phonetic properties of the novel reduplicated words and nonreduplicated words are summarized in Tables 1 and 2. Phonotactic probabilities were calculated using the Child Mental Lexicon (Storkel & Hoover, 2010), summing the position-specific probabilities of segments and biphones in each of the four reduplicated words and four nonreduplicated words. Neighborhood frequency was calculated as the sum of the log frequencies for the neighbors. The phonetic comparisons were based on the duration, amplitude, and minimum/maximum fundamental frequency (F0) of the two tokens of each experimental word in the familiarization trials and test trials. No statistical difference was found between the reduplicated and nonreduplicated words in terms of their phonotactic probabilities [all $ts(6) < 1$], neighborhood properties [all $ts(6) < 1$] or phonetic measurements [all $ts(14) < 1.3$].

<Insert Table 1 here>

<Insert Table 2 here>

The visual stimuli consisted of high-definition color photographs of four familiar and two novel items, averaging approximately 20 cm in diameter when displayed on the monitor used in the experiment. The familiar items were a dog, a rabbit, a balloon, and an apple, corresponding to the familiar words used in the task. The novel items were images of stuffed toys, shown in Figure 1.

<Insert Figure 1 here>

Apparatus. The experiment was carried out in a dimly-lit sound-attenuated room, equipped with a 47-inch flat-screen monitor to present the visual stimuli, a set of loudspeakers, and a remote-controlled digital video camera installed underneath the monitor to record the eye-gaze of the infant. The audio and visual stimuli were controlled by the Habit X 1.0 programme (Cohen, Atkinson, & Chaput, 2004).

Procedure. Before the experiment, the caregiver of the participant completed a short inventory of noun vocabulary in order to report the infant's familiarity with the real words used in the study. During the experiment, the infant sat on the lap of the caregiver, who was seated approximately 1.5 meters away from the display monitor. The caregiver was instructed to listen to masking music played over headphones and not to speak to the infant or point at the visual images during the experiment.

The experiment consisted of 10 familiarization trials and 12 test trials, presented in that order. All trials lasted 7 seconds and were interspersed with an attention-getting video sequence that showed colorful moving bubbles with a soundtrack of children's laughter. Trials were initiated by an experimenter in an adjacent control room, once they had confirmed through a computer monitor that the infant was visually fixated to the attention-getting sequence. The experiment lasted approximately 4.5 minutes.

During the familiarization trials, each participant was exposed to all four real word-image combinations, but only to one reduplicated word (*neenee*, *foofoo*, *bobo*, or *laylay*) and one nonreduplicated word (*neefoo*, *foonee*, *bolay*, or *laybo*), each paired with one of the two novel objects. The combination of the reduplicated word and nonreduplicated word for each participant was constrained such that the two novel words must not share a syllable. Thus, an infant who was given *neenee* as a reduplicated word was given either *bolay* or *laybo* as a nonreduplicated word, but not *neefoo* or *foonee*. This generated eight possible experimental combinations of the reduplicated and nonreduplicated words (*neenee* – *bolay*, *neenee* – *laybo*, *bobo* – *neefoo*, etc.). Each of these eight combinations was assigned to three participants for counterbalancing. The combination of the novel words and the objects was also counterbalanced across participants such that, for example, half the participants who heard *neenee* heard it paired with one of the novel toys, but the other half heard it paired with the other novel toy.

A familiarization trial began with a single image appearing at the central bottom of the screen. This image was one of the familiar objects (i.e., dog, rabbit, apple, and balloon) or the unfamiliar objects (i.e., stuffed toys described in the Materials section). Over the course of the 7 s trial, the image then moved up and down once along the horizontal center at a steady rate. Two seconds after the trial began, the first sentence was played with the test word matching the image (e.g., *Look! A doggy.*). The second carrier phrase began 4 s into the trial (e.g., *Yes, that's a doggy!*). The ten familiarization trials included one trial for each of the four familiar words (i.e., *apple*, *balloon*, *doggy*, and *bunny*), three trials for a reduplicated novel word (e.g., *neenee*) and three trials for a nonreduplicated novel word (e.g., *bolay*). Each of the three novel word trials was embedded in a different carrier sentence. The carrier sentence sets for the real words were rotated across words and participants. The order of trials was pseudo-randomized, with real words always presented in trials 1, 2, 6, and 8. The first

two trials with real words were intended to highlight the nature of the task to the participants and ensure that they engaged with the task. Real words were also positioned in Trials 6 and 8 to break up the sequence of novel words evenly and to keep the participants' attention. The order of the novel words was otherwise randomized, except that the same word was not played more than twice in a row.

Each test trial began with two images appearing side by side at the bottom of the screen, separated by about 52 cm. The images were drawn from the same six photographs used in the familiarization phase, in three fixed pairs. The familiar objects formed two pairs matched for animacy (i.e., the dog always appeared with the rabbit, and the apple always appeared with the balloon), and the two unfamiliar objects formed another pair. In each trial, one member of the pair was the target object matching the test word, and the other object served as the competitor. For example, when *doggy* was the test word, the target object was a dog and the competitor was a rabbit. When a nonword was the test item, the matching novel stuffed toy was the target object and the competitor was the other novel toy. During the first 2 s of the test trial, the images simultaneously moved up toward the vertical center of the screen and stayed in that position. At this point the first sentence was played with the test word matching the target image (e.g., *Where's the neenee?*). The onset of the test word (e.g., *neenee*) was aligned at the 2.5 s point of the trial. At the 5 s point, the competitor disappeared and the target began to rotate along the center of the image, tilting first outwards and then inwards, changing its direction cyclically when it reached a 30 degree rotation. This 'dancing' sequence at the end of each test trial was used to create a pragmatically natural context for labeling, highlighting the referential nature of the task (Halberda, 2003; Killing & Bishop, 2008). The second sentence (e.g., *That's the neenee!*) accompanied this animation sequence, to provide further opportunities for learning in the second half of the test trial. The test measure was therefore only obtained from the instance of the target word in the first

(question) sentence (e.g., *Where's the neenee?*). The timeline of the test trial is illustrated in Figure 2.

<Insert Figure 2 here>

There were 12 test trials in total: one trial for each of the four real words and four trials for each novel word. The order of these trials was pseudo-randomized, with several constraints. Trials 1 and 2 were always with real words so that the nature of the task could be established in the initial trials. The other two real word trials were placed in Trials 7 and 10 to break up the sequence of novel words evenly and also to ensure that there were exactly four novel word trials in the first half (Trials 3, 4, 5 and 6) and four novel word trials in the second half (Trials 8, 9, 11 and 12). The target side was never the same on more than two consecutive trials. Each novel word had its target image shown once on one side and once on the other side in each half of the experiment, and also appeared with one carrier sentence set (e.g., *Where's the neenee? That's the neenee.*) in the first half of the experiment, and the other set (e.g., *Where's the neenee? There's the neenee.*) in the second half. The same novel word was never a target on more than two consecutive trials, nor on the first two novel word trials (i.e., Trials 3 and 4).

Coding. The infants' eye gaze throughout the experiment was video-recorded in MPEG format. The videos were divided into two halves, each of which was analyzed by one of two trained coders, who stepped through the video-recording using custom-made software, noting for each 40 ms frame whether the infant was looking at the left image, right image, or neither. Inter-coder reliability was assessed using all the trials from three randomly selected participants. Mean agreement rate between the two coders was 98.42% (range: 97.41% - 99.35%), and mean Cohen's Kappa was .974 (range: .957 - .988).

Results and Discussion

The critical phase of the test trial was the first 5 seconds, before the competitor image disappeared and the target image began to move. During this phase, if the infant did not look at either object at least 60% of the time, the entire trial was excluded from the analysis. By this criterion, 14.9% of trials (43 out of 288) were discarded. As noted above, all infants analyzed had at least one trial that met this condition for each word type.

The index used to measure preferential looking was the target look proportion, calculated as the proportion of looks to the target image over the total amount of looks to either the target or competitor image. For each trial, this measure was taken for the baseline phase (the 2 s period leading up to the onset of the test word) and post-naming phase (the 2 s period beginning from 360 ms after the onset of the test word). A change score was then calculated by subtracting the baseline target look proportion from the post-naming target look proportion for each trial, and then averaged across infants. A change score above chance level would indicate preferential looking toward the correct word-object matching.

The first analysis asked whether the change scores varied during the course of the experiment. As the test trials in our design included a familiarization component at the end, performance might have improved over trials. At the same time, the performance in later trials might have been negatively affected by fatigue. In order to examine such trial order effects, mean change scores were calculated for the first, second, third, and last trial for each of the three word types (i.e., real words, reduplicated words, and nonreduplicated words). The results are summarized in Figure 3. Although there is an apparent dip in the change score for the second real-word trial, none of the word types exhibited an overall improvement or decline over time. This was confirmed by a two-way repeated-measures ANOVA with Word Type and Order¹. The main effect of Word Type ($F(2, 46) = 4.74, p = .013, \eta^2 = 0.041$) was significant, but neither the main effect of Order ($F(3, 69) = 0.72, p = .543, \eta^2 = 0.007$) nor the

¹ Missing values due to the exclusion criteria described above were replaced by the Word Type x Order means calculated from included trials.

interaction between Word Type and Order ($F(6, 138) = 0.56, p = .765, \eta^2 = 0.011$) was significant. Based on these results, we collapsed the data across trials for subsequent analyses.

<Insert Figure 3 here>

Next, we compared the participants' overall preferential looking between word types. A one-way repeated-measures ANOVA revealed a significant effect of Word Type on the change score ($F(2, 46) = 5.00, p = .011, \eta^2 = 0.120$). A planned comparison showed that reduplicated words had a higher change score ($M = 0.11, sd = 0.22$) than nonreduplicated words ($M = -0.01, sd = 0.18, t(23) = 1.96, p = .031$, one-tailed, $d = 0.40$). One-sample t-tests against chance showed that target look proportion was significantly higher in the post-naming phase than in the baseline phase for real words (change score $M = 0.16, sd = 0.20, t(23) = 3.97, p < .001, d = 0.81$) and reduplicated words ($M = 0.10, sd = 0.22, t(23) = 2.35, p = .028, d = 0.48$), but not for nonreduplicated words ($M = -0.01, sd = 0.18, t(23) = 0.36, p = .725, d = 0.07$). This contrast between real/reduplicated words vs. nonreduplicated words could not be attributed to a target fixation difference in the baseline phase, which was not statistically reliable ($F(2, 46) = 1.29, p = .283, \eta^2 = 0.044$). Specifically, there was no difference in baseline target look proportions between reduplicated ($M = 0.50, sd = 0.16$) and nonreduplicated words ($M = 0.53, sd = 0.15, t(23) = 0.984, p = .335, d = 0.20$).

<Insert Figure 4 here>

In the proportion-to-target analyses above, we used a window of analysis beginning 360 ms after the onset of the test word and extending for 2 s, following the commonplace practice in this type of experiment with 18-month-olds (Fernald, Pinto, & Swingley, 1998; Fernald, Zangl, Portillo, & Marchman, 2008). The timing of this window is known to best capture the variability in eye movement responses, but it has also been noted that 18-month-olds need more time to reach asymptote in their response to novel words (Swingley, 2011). In

order to examine the patterns of target fixation over time following the presentation of the test word, we produced a timecourse plot displaying continuous changes in target fixation (Figure 5). As the plot shows, a shift toward the target object can be observed for all three word types, including nonreduplicated words. However, there is a difference in the timing of the shift, with target fixation peaking some time between 3,500 and 4,000 ms (i.e., 1,000-1,500 ms post-stimulus-onset) for real words and reduplicated words, but not until the 4,000 ms point for nonreduplicated words. To test whether a reliable target fixation can be observed for all word types if we shifted the window of analysis for longer latency, we calculated the mean proportions of target looks between 3,500 and 5,000 ms (i.e., 1,000-2,500 ms post-stimulus-onset). These were significantly above chance not only for real words ($M = 0.65$, $t(23) = 4.68$, $p < .001$, $d = 0.95$) and reduplicated words ($M = 0.60$, $t(23) = 2.29$, $p = .031$, $d = 0.48$), but also for nonreduplicated words ($M = 0.57$, $t(23) = 2.08$, $p = .049$, $d = 0.42$). However, when the target look proportions in the late window were compared with those in the baseline (i.e., 500-2,500 ms post-trial onset), the difference was significant for real words ($M = 0.18$, $t(23) = 4.12$, $p < .001$, $d = 0.84$) and close to the alpha-level for reduplicated words ($M = 0.11$, $t(23) = 2.01$, $p = .056$, $d = 0.41$), but not significant for nonreduplicated words ($M = 0.02$, $t(23) = 0.54$, $p = .593$, $d = 0.11$). Thus, while the infants were responding to both the reduplicated and nonreduplicated words in this late window of analysis, the response to nonreduplicated words was weaker than that to reduplicated words.

<Insert Figure 5 here>

In sum, infants showed short-latency fixation toward the target object in response to a familiar real word or a novel reduplicated word, but not in response to a novel nonreduplicated word. There were some signs of late-latency target fixation in response to a novel nonreduplicated word as well, but the effect was not as reliable as that for familiar real

words or reduplicated words. These results indicate that infants' learning of novel word-object mapping is more robust with reduplicated words than with nonreduplicated words.

Corpus Analysis

The results of our experiment show that reduplicated words are privileged in early word learning when tested with English-exposed 18-month-olds. However, it is not clear whether this experimental effect has roots in an experience-independent processing bias or an experience-dependent readiness to learn reduplicated words. Given the known tendency for register-specific infant-directed words to feature reduplication (e.g., *choochoo*, *wee-wee*), it is possible that the linguistic input to English-exposed infants contains frequent syllable reduplication, leading them to an acquired bias toward reduplicated words. To test this possibility, we analyzed the occurrence frequencies of reduplicated words as well as repetition of adjacent syllables in naturalistic corpora of maternal speech addressed to infants.

Reduplicated Words in the Input

As a first step in the corpus analysis, we examined the relative lexical frequency of disyllabic words consisting of two identical syllables (e.g., *choochoo*) in a typical sample of infant-directed speech. The analysis was performed on the longitudinal spontaneous speech data in the Brent Corpus (Brent & Siskind, 2001), accessed through the Child Language Data Exchange System (CHILDES; MacWhinney, 2000). The corpus contains approximately 100 hours of bi-weekly recordings of 16 English-speaking mothers interacting with their children (6 females, 10 males), who were 8 to 9 months old at the beginning of the data collection and 14 to 15 months at the end. In total, 9,397 word types and 491,779 word tokens were identified in the maternal speech addressed to their child.

As expected, almost all the reduplicated words found in the corpus were lexical items that would be regarded as those specific to the register of infant-directed speech (i.e., 'baby-

talk words'). The ten most frequently-used reduplicated words were *bye-bye*, *night-night*, *peepee*, *weewee*, *poopoo*, *booboo*, *yum-yum*, *choochoo*, *woof-woof*, and *doodoo*. However, the occurrences of these words do not make disyllabic reduplication prevalent in the overall composition of the lexical input to infants. On average, the proportion of disyllabic reduplicated words in all the infant-directed words produced by the 16 mothers was merely 0.7 percent by type count and 0.6 percent by token count. Reduplicated words accounted for only 2.1 percent (type) or 3.1 percent (token) of all disyllabic words, the remainder of disyllabic words being those with no syllable repetition (e.g., *mommy*, *kitty*, *doggy*, *bunny*). Even when the comparison was made only for nouns (which may be preferentially learned by 18-month-olds), only 1.0 percent (type) or 0.7 percent (token) of all disyllabic nouns were reduplicated.

These observations show that infants' vocabulary input is not predominated by reduplicated words by any measure. As a matter of fact, reduplicated words are much less frequent than nonreduplicated disyllabic words. Thus, we can at least rule out the possibility that the outcome of the word learning experiment was induced simply by infants' exposure to a vocabulary in which the majority of words (or disyllabic words/nouns) are reduplicated.

Phonological Analysis of Reduplication in the Input

While the analysis presented above demonstrates that reduplicated words do not predominate the lexical input to English-learning infants, it does not offer any information about how frequently the overall phonological input contains adjacent repetition of syllables. Because infants are capable of extracting transitional probabilities between adjacent syllables in the input (Saffran, Aslin, & Newport, 1996; Aslin, Saffran, & Newport, 1998), they may be tracking the probability of any syllable being immediately followed by an identical syllable, whether or not those syllables are contained in a single word. If this observed probability is higher than what one would expect from chance occurrences of adjacent

syllable repetition, infants may come to see adjacent repetition of syllables as a relatively frequent pattern in the phonological input of their ambient language.

To investigate this possibility, we examined the occurrences of exact adjacent repetition of syllables in the Bernstein Ratner corpus (Bernstein Ratner, 1984), which consists of transcribed speech of nine mothers speaking to their children, who were 13 to 21 months of age (mean = 18 months). This corpus is smaller than the Brent corpus that was used for the analysis of reduplicated words presented earlier, but it was selected here because of the availability of a phonologically-represented derived corpus prepared by Brent and Cartright (1996) and contributed to CHILDES by Sharon Goldwater. This version of the corpus contains 1,179 word types and 38,693 word tokens in total.

To calculate the observed frequency of reduplication, we first separated the phonemic transcripts into arrays of syllables and took the proportion of adjacent repetitions within the number of syllable transitions in each mother's speech.² We did not include sequences crossing utterance boundaries, as we assumed infants compute such statistics within utterances. As a baseline of comparison, we ran a Monte Carlo simulation of the chance likelihood of adjacent syllable repetition by combining two syllables randomly drawn from the entire set of syllables recorded for each mother. The probability of such combination being identical syllables was calculated by dividing the number of matches by the total number of syllable transitions, and repeating this process 10,000 times per mother.

The results are presented in Table 3 and Figure 6. The Z-scores in Table 3 show how much the observed frequencies deviate from the estimated means of the simulations. A positive score would indicate an observed frequency that is higher than expected by chance and a negative score would indicate an observed frequency that is lower than expected by chance. For most mothers, the frequency of syllable repetition in the input sample was either

² The syllabified corpus is available by request from the first author.

significantly lower than the simulation means, or no more or less than chance level. As a whole, the observed frequencies in the nine mothers were lower than the simulation means ($t(8) = 2.69, p = .028, d = 0.90$). There was one exception to this pattern (Mother 7), where the observed frequency was significantly higher than the level predicted by random combinations. Upon close inspection, the transcripts from this dyad revealed many situations that invited the repeated use of the words/phrases *night-night* and *bye-bye*, resulting in a higher count of syllable repetition than other dyads. This is consistent with the observation made by Daland (2013) that phonological frequencies in infant-directed speech exhibit a ‘bursty’ distribution that is driven primarily by discourse topic. Such occasional deviations are therefore likely to be lost in the large cross-situational variation in the phonological input to the learner.

<Insert Table 3 here>

<Insert Figure 6 here>

To summarize, the analysis of syllable transitions in maternal speech data presented here indicates that the actual frequency of adjacent syllable repetition in the input is generally lower than would be expected by chance. Taken together with the analysis of reduplicated words in infant-directed speech, these findings show that infants’ bias in learning reduplicated words relative to nonreduplicated words is unlikely to be a product of a frequency effect from the ambient linguistic input.

General Discussion

The purpose of this study was to examine the potential role of a bias for repetition in early word learning. As a first step, we tested whether reduplicated words are easier to learn by comparing infants’ novel word-object mappings involving a reduplicated $C_1V_1C_1V_1$ word versus a nonreduplicated $C_1V_1C_2V_2$ word. The results showed that under the experimental

conditions set in our study, 18-month-olds were better at learning a reduplicated word than a nonreduplicated word. As the novel words were controlled for their phonetic and phonotactic characteristics, the difference in the results is best explained as a reduplication bias in early word learning. Thus, infants' disposition to learn repetitions in the input structure is not limited to the pattern generalization of type repetitions, but also extendable to the lexical learning of token repetitions.

The question we asked next was whether this asymmetry in learning word-object mappings could be understood as evidence for an inherent disposition to preferentially attend to structures with repetition (e.g., Endress et al., 2009; Gervain & Werker, 2008). As mentioned in the introduction, there is a caveat in drawing such a conclusion from these experimental results: the infants' preference might not reflect an experience-independent bias, but rather a readiness to learn word forms with patterns that are more frequent in the input. As our corpus analysis showed, however, this is not the case. A typical sample of infant-directed speech around or prior to 18 months is not predominated by disyllabic words with reduplication, nor does it have more repetitions of identical syllables than expected by chance. Thus, the observed reduplication advantage cannot be attributed to the prominence of reduplicated words or repeated syllables in the learner's linguistic environment.

One might nevertheless argue that this effect still originates in the input pattern because reduplicated words and syllables are statistically *infrequent*, rather than frequent, in the overall input to English-learning infants. This possibility deserves consideration given that infants may rapidly generalize patterns such as adjacent repetition (e.g., *leledi* → *kokoba*) when the relevant strings are extremely infrequent in the input (Gerken et al., 2015). In such circumstances, even a single example of the rare pattern can alert infants to update their current hypothesis of likely linguistic structures. Although this interpretation might hold true for rule generalization or the learning of type repetitions (e.g., *leledi* → AAB), it is unlikely

to apply to the domain of lexical acquisition or token repetitions (e.g., learning the lexical item *lele* as opposed to *ledi*), as, to our knowledge, all phonological biases in word segmentation and learning that are sensitive to input patterns are in the direction toward a predominant, not rare, structure in the input, as seen in the bias for labial-coronal sequences (Gonzalez-Gomez et al., 2014 and Højen & Nazzi, 2016), predominant stress patterns (Curtin et al., 2005 and Jusczyk et al., 1993), and higher phonotactic probabilities (Storkel, 2009). The overall evidence is therefore in favor of the interpretation that the bias toward learning reduplicated words is independent of the relevant patterns in infants' linguistic input.

Interestingly, this learning bias toward repetition does not appear to be reflected in the general patterns in natural languages. In fact, mature languages show the opposite trend. Although full reduplication is frequently employed as a morphological process (Hurch & Mattes, 2009; Moravcsik, 1978), it is dispreferred in word roots and stems across languages. Even repetition of consonants within a word is avoided to some degree (Bonatti, Peña, Nespor, & Mehler, 2007). For example, many languages have disproportionately fewer roots containing consonants drawn from the same place of articulation than those with place-differentiated consonants (MacNeilage et al., 2000; Monaghan & Zuidema, 2015; Rousset, 2004; Pozdniakov & Segerer, 2007). The tendency for lexicons to avoid similar-sound repetition has led phonologists to propose the Obligatory Contour Principle (OCP; Leben, 1973; McCarthy, 1986), a constraint that bans morpheme-internal sequences of elements (e.g., consonants, tones) that share a phonological feature in underlying representations. While its precise formulation and scope have been a matter of debate, the OCP captures the observation that many phonological patterns across languages can be explained in terms of avoidance of consecutive identical features. Furthermore, recent experimental work indicates that repetition avoidance akin to the OCP may also affect adult speakers' word segmentation. When exposed to an auditorily presented sequence of syllables such as /po/, /ma/, /tu/, and

/ne/, adult speakers of Dutch are more likely to remember a string that does not contain consonant repetition between adjacent syllables (e.g., /potupa/) than one that does (e.g., /popatu/) (Boll-Avetisyan & Kager, 2014).

Our finding that infants are better at learning reduplicated words than nonreduplicated words is therefore diametrically opposed to the generalizations drawn from typological, phonological and psycholinguistic research that the adult lexicon tends to avoid words containing repeated syllables and consonants. This disconnect between a bias in language learning and generalizations in mature linguistic systems seems to be at odds with the notion that typological universals and fundamental characteristics of human language are manifestations of learning biases that constrain language acquisition (e.g., Bever, 1970; Chater & Christiansen, 2010; Moreton, 2008; although see Blevins, 2004; Evans & Levinson, 2009 for opposing views).

One might still suggest that our results are in fact consistent with OCP-like effects if we assume that the infants did avoid within-word repetition and parsed the reduplicated word as two tokens of a monosyllabic word rather than one disyllabic word (e.g., “*nee ... nee*” rather than “*neenee*”). If so, the infants were exposed to the hypothetical monosyllabic word (e.g., *nee*) twice as frequently as the nonreduplicated disyllabic word (e.g., *bolay*), leading to the differential responses to the reduplicated versus nonreduplicated items. As we did not test our infants on their recognition of just one of the repeated syllables (e.g., *nee*), we do not have evidence that speaks directly to this possibility. Nonetheless, there are reasons to believe that this interpretation is highly unlikely. Both types of disyllabic words in our experiment – reduplicated as well as nonreduplicated – were spoken with word-level stress and a pitch accent only on the initial syllable, and were prosodically coherent as a disyllabic word form. It is well established that English-learning infants show marked preference for segmenting

words with strong/weak stress patterns from running speech (Jusczyk et al., 1993, 1999), a preference that must have guided our participants' learning of both types of words.

Why would then infant word learning not be subjected to the OCP? One possibility is that OCP-like effects in adult language could be due to perceptual or processing factors: repeated sounds can be misperceived as a single sound (Boersma, 2000) or they may be difficult to process because the encoding of the second instance of the repeated sound is inhibited by the first (Frisch, 2004). However, these factors may not weigh heavily in infants' language processing due to the slow speech rate in infant-directed speech (see MacNeilage et al., 2000 for a related argument for the disconnect between the OCP and early speech production). Alternatively, repetition avoidance may be a principle that strictly applies to the phonological organization of the lexicon (Boll-Avetisyan & Kager, 2014). If this is the case, then speakers acquire the tendency to avoid repetition in lexical processing only by making phonotactic generalizations over the vocabulary items of the language. Infants, however, may have not learned enough words from the ambient language to detect patterns of repetition avoidance.

Relatedly, adult lexicons may not reflect the putative learning bias for repeated structures because word-internal repetitions would restrict the combinatorial possibilities for generating phonologically distinct lexical items. In the adult lexicon, therefore, the pressure to have more learnable word forms may be counteracted by the demand to optimize expressivity (i.e., the ability to encode semantic differences with as little redundancy as possible). In contrast, there is less pressure to maximize phonological differentiation in the lexicon addressed to infants because of its relatively small size, and that may explain why we commonly find reduplication in register-specific lexical items (e.g., *mama* 'food' in Gilyak, *toto* 'bath' in Marathi, *nene* 'baby' in Spanish; Ferguson, 1964, 1978). As our corpus analysis shows, the addition of these words still does not make reduplicated words a predominant form in the overall lexical input. However, even a few such words may serve as an entry

point into lexical development, or what Caselli et al. (1995) have termed a ‘starter set’ for vocabulary acquisition.

While these ideas are speculative, they are in line with other analogous cases where a learning bias is not mirrored in the general properties of natural language but is nonetheless accommodated in infant-directed vocabulary. For example, infants and children are generally more likely to learn words with sound-meaning mappings that are internally systematic (Monaghan, Shillcock, Christiansen, & Kirby, 2014) or iconic in nature (Imai, Kita, Nagumo & Okada, 2008; Maurer, Pathman, & Mondloch, 2006). This is a bias that runs against the largely arbitrary relationship between the phonology and semantics of mature lexicons. Yet, infant-directed vocabulary appears to accommodate the learning bias by containing more onomatopoeic items that have systematic sound-meaning mappings (Caselli et al., 1995). Similarly, learners are better at segmenting a set of novel words in running speech if the words have phonologically invariant edges – all ending in /tjə/, for example (Kempe, Brooks, & Gillis, 2005; Kempe, Brooks, Gillis, & Samson, 2007). Again, this learning bias counters the ample phonological variance in word edges observed in natural languages, but accords with the tendency for baby-talk words to have reduced edge variance due to the use of diminutives (e.g., *doggy*, *daddy*, *froggy*). The advantage for reduplicated words observed in our study can also be portrayed as an example of such a bias, which the mature system has incorporated only in the service of bootstrapping into lexical learning.

Returning to more empirical issues, an important question left unanswered is where exactly the effects of reduplication emerge during the process of lexical learning. Repeated syllables may be more easily detected in running speech, making them more likely to be successfully segmented as a unit. Reduplicated words may allow more efficient phonological encoding in the lexicon because of their high level of redundancy. The salience of adjacent repetition may also mean that reduplicated words have a higher level of lexical activation,

which allows faster recognition. Any of these possibilities can explain the results of our experiment, because the task involved all three aspects: segmenting the novel words from the carrier sentences, phonologically encoding the word form as a label for the unfamiliar object, and recognizing the learned word form during the test phase. Identifying the locus of the reduplication advantage in lexical learning is a task we aim to explore in future work.

Another question we were not able to fully address in this study is the nature of children's interpretation of the novel labels. The real words in the task were likely to be understood as basic level categories (i.e., dog, rabbit, apple, balloon). It is not clear, however, what type of semantic interpretation was assigned to the referents of the novel words. Although the novel words in the experiment were intended as labels for a general category ("Look, *a* neenee", "Where's *the* neenee?"), we cannot rule out the possibility they were understood as individual names of the animate toys. It thus remains to be seen whether the reduplication advantage discovered in this study is limited to a certain type of referents, such as single individuals, or is extendable to a broader range of semantic categories. This can be investigated by manipulating the level of object-kind taxonomic mapping infants are likely to learn based on the exemplars they are exposed to during experimental familiarization (see Xu & Tenenbaum, 2007).

To conclude, we have demonstrated that, all other things being equal, 18-month-old infants exposed to English are more adept at learning reduplicated words than nonreduplicated words. This advantage for reduplicated words is not a reflection of a frequent pattern in the lexical or phonological input to infants; instead it is more likely to come from an experience-independent phonological bias in word learning. Thus, repeated elements are privileged not only in pattern generalization, but also in token detection and learning, particularly in the context of lexical development. We have seen that such learning biases may not shape the structure of natural language lexicons, but may give rise to a

compartmentalized adaptation in the form of infant-directed vocabulary. If so, learning biases and generalizations in mature linguistic systems are not necessarily coupled, but are likely to have a more complex relationship which emerges from the interaction between the demand to optimize the internal structure of the system and the demand to make the system learnable.

References

- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9, 321-324.
- Bernstein Ratner, N. (1984). Patterns of vowel modification in mother-child speech. *Journal of Child Language*, 11, 557-578.
- Bever, T. (1970). The cognitive basis for linguistic universals. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 227-360). New York, NY: Wiley & Sons.
- Bird, E. K. R., Chapman, R. S., & Schwartz, S. E. (2004). Fast mapping of words and story recall by individuals with Down syndrome. *Journal of Speech, Language, and Hearing Research*, 47, 1286-1300.
- Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.
- Boersma, P. (2000). *The OCP in the perception grammar*. Rutgers Optimality Archive, ROA-435, <http://roa.rutgers.edu/article/view/445>.
- Boll-Avetisyan, N., & Kager, R. (2014). OCP-PLACE in speech segmentation. *Language and Speech*, 57, 394-421.
- Bonatti, L., Peña, M., Nespor, M. & Mehler, J. (2007). On consonants, vowels, chickens, and eggs. *Psychological Science*, 8, 924-925.
- Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, 61, 93-125.
- Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81, B33-B44.
- Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, 10, 159-199.

Chater, N., & Christiansen, M. H. (2010). Language acquisition meets language evolution.

Cognitive Science, 34, 1131-1157.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT press.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2004). *Habit X: A new program for*

obtaining and organizing data in infant perception and cognition studies (Version 1.0).

Austin: University of Texas.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: statistical learning of

nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, 30, 1119.

Curtin, S., Mintz, T. H., & Christiansen, M. H. (2005). Stress changes the representational

landscape: Evidence from word segmentation. *Cognition*, 96, 233-262.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior*

Research Methods, Instruments, & Computers, 28, 125-127.

de Vries, M. H., Petersson, K. M., Geukes, S., Zwitserlood, P., & Christiansen, M. H. (2012).

Processing multiple non-adjacent dependencies: Evidence from sequence learning.

Philosophical Transactions of the Royal Society B: Biological Sciences, 367, 2065-

2076.

Endress, A., Dehaene-Lambertz, G., & Mehler, J. (2007). Perceptual constraints and the

learnability of simple grammars. *Cognition*, 105, 577-614.

Endress, A. D., Nespore, M., & Mehler, J. (2009). Perceptual and memory constraints on

language acquisition. *Trends in Cognitive Sciences*, 13, 348-353.

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of

algebraic rules. *Journal of Experimental Psychology: General*, 134, 406.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity

and its importance for cognitive science. *Behavioral and Brain Sciences*, 32, 429-448.

- Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, 407, 630-633.
- Feldman, J. (2003). The simplicity principle in human concept learning. *Current Directions in Psychological Science*, 12, 227-232.
- Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist*, 66, 103-114.
- Ferguson, C. A. (1978). Talking to children: A search for universals. In J. Greenberg, C. A. Ferguson & E. A. Moravcsik (Eds.), *Universals of human language. Volume 1. Method and theory* (pp. 203-224). Palo Alto, CA: Stanford University Press.
- Fernald, A., Pinto, J. P., Swingle, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, 9, 228-231.
- Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. In I. A. Sekarina, E. M. Fernández, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 113-132). Amsterdam: John Benjamins.
- Frisch, S. A. (2004). Language processing and segmental OCP effects. In B. Hayes, R. M. Kirchner, & D. Steriade (Eds.) *Phonetically based Phonology* (pp. 346-371). Cambridge: Cambridge University Press.
- Gerken, L. (2010). Infants use rational decision criteria for choosing among models of their input. *Cognition*, 115, 362-366.
- Gerken, L., Dawson, C., Chatila, R., & Tenenbaum, J. (2015). Surprise! Infants consider possible bases of generalization for a single input example. *Developmental Science*, 18, 80-89.

- Gervain, J., Berent, I., & Werker, J. F. (2012). Binding at birth: The newborn brain detects identity relations and sequential position in speech. *Journal of Cognitive Neuroscience*, 24, 564-574.
- Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences*, 105, 14222-14227.
- Gervain, J., & Werker, J. F. (2008). How infant speech perception contributes to language acquisition. *Language and Linguistics Compass*, 2, 1149-1170.
- Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70, 109-135.
- Gomez, R. L., Gerken, L., & Schvaneveldt, R. W. (2000). The basis of transfer in artificial grammar learning. *Memory & Cognition*, 28, 253-263.
- Gonzalez-Gomez, N., Hayashi, A., Tsuji, S., Mazuka, R., & Nazzi, T. (2014). The role of the input on the development of the LC bias: A crosslinguistic comparison. *Cognition*, 132, 301-311.
- Gonzalez-Gomez, N., & Nazzi, T. (2012). Acquisition of nonadjacent phonological dependencies in the native language during the first year of life. *Infancy*, 17, 498-524.
- Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87, B23-B34.
- Houston-Price, C., Plunkett, K. I. M., & Harris, P. (2005). 'Word-learning wizardry' at 1;6. *Journal of Child Language*, 32, 175-189.
- Højen, A., & Nazzi, T. (2016). Vowel bias in Danish word-learning: Processing biases are language-specific. *Developmental Science* 19(1), 41-49.
- Hurch, B., & Mattes, V. (2009). Introduction: Diachrony and productivity of reduplication. *Morphology*, 19, 107-112.
- Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109, 54-65.

- Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, 64, 675-687.
- Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39, 159-207.
- Kempe, V., Brooks, P. J., & Gillis, S. (2005). Diminutives in child-directed speech supplement metric with distributional word segmentation cues. *Psychonomic Bulletin & Review*, 12, 145-151.
- Kempe, V., Brooks, P. J., Gillis, S., & Samson, G. (2007). Diminutives facilitate word segmentation in natural speech: Cross-linguistic evidence. *Memory & Cognition*, 35, 762-773.
- Killing, S. E., & Bishop, D. V. (2008). Move it! Visual feedback enhances validity of preferential looking as a measure of individual differences in vocabulary in toddlers. *Developmental Science*, 11, 525-530.
- Leben, W. R. (1973). *Suprasegmental phonology* (Doctoral dissertation), Massachusetts Institute of Technology, Cambridge, Massachusetts.
- McCarthy, J. J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry*, 17, 207-263.
- MacNeilage, P. F., Davis, B. L., Kinney, A., & Matyear, C. L. (2000). The motor core of speech: A comparison of serial organization patterns in infants and languages. *Child Development*, 71, 153-163.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed). Mahwah, NJ: Lawrence Erlbaum.
- Mani, N., & Plunkett, K. (2010). Twelve-month-olds know their *cups* from their *keps* and *tups*. *Infancy*, 15, 445-470.

- Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283, 77-80.
- Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, 9, 316-322.
- Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20130299.
- Monaghan, P., & Zuidema, W. (2015). General purpose cognitive processing constraints and phonotactic properties of the vocabulary. Paper presented at the Workshop on the Evolution and Phonetic Capabilities: Causes, Constraints and Consequences. Retrieved from http://www.lancaster.ac.uk/staff/monaghan/papers/monaghan_zuidema_15_icphs.pdf.
- Moravcsik, E. A. (1978). Reduplicative constructions. In J. H. Greenberg (Ed.), *Universals of human language (vol. 3): Word structure* (pp. 297-334). Stanford, MA: Stanford University Press.
- Moreton, E. (2008). Analytic bias and phonological typology. *Phonology*, 25, 83-127.
- Moreton, E. (2012). Inter-and intra-dimensional dependencies in implicit phonotactic learning. *Journal of Memory and Language*, 67, 165-183.
- Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, 98, 13-30.
- Newport, E. L. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Language Sciences*, 10, 147-172.
- Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 80.

Pozdniakov, K., & Segerer, G. (2007). Similar place avoidance: A statistical universal.

Linguistic Typology, 11, 307-348.

Rousset, I (2004). *Structures syllabiques et lexicales des langues du monde: Données, typologies, tendances universelles et contraintes substantielles*. PhD dissertation.

Université Stendhal, Grenoble, France.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, 47, 172-196.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, 39, 484.

Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development*, 69, 309-320.

Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of Child Language*, 36, 291-321.

Storkel, H. L., & Hoover, J. R. (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods*, 42, 497-506.

Swingle, D. (2011). The looking-while-listening procedure. In E. Hoff (Ed.), *Research method in child language: A practical guide* (pp. 29-42). Oxford: Blackwell.

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology*, 44, 929-938.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, 97, B25-B34.

- Tunney, R. J., & Altmann, G. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 614-639.
- Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, 34, 1289-1309.
- Wexler, K., & Culicover, P. (1980). *Formal principles of language acquisition*. Cambridge, MA: MIT Press.
- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, 114, 245-272.

Tables

Table 1

Means (and Standard Deviations) of the Phonotactic Probabilities and Neighborhood Properties of the Novel Words

	Reduplicated	Nonreduplicated
Positional segment probability (sum)	.1774 (.061)	.1774 (.087)
Biphone probability (sum)	.0076 (.0057)	.0087 (.0048)
Number of neighbors	1.50 (2.38)	1.00 (1.41)
Neighborhood log frequency (sum)	2.18 (2.93)	2.19 (2.56)

Table 2

Means (and Standard Deviations) of the Phonetic Properties of the Novel Words

	Familiarization		Test	
	Reduplicated	Nonreduplicated	Reduplicated	Nonreduplicated
Duration (ms)	779 (91)	759 (86)	716 (106)	728 (92)
Intensity (dB)	75.6 (1.1)	75.0 (1.3)	76.1 (1.4)	76.5 (0.8)
Min F0 (Hz)	135 (3)	135 (3)	133 (4)	135 (3)
Max F0 (Hz)	393 (22)	390 (18)	312 (65)	318 (60)

Table 3

Observed Frequencies of Syllable Repetition and Means (and Standard Deviations) of Simulated Frequency Distribution

Child	Observed frequency	Simulation mean (SD)	Z-score	<i>p</i>
1	.0059	.0070 (.0013)	-0.84	.401
2	.0002	.0087 (.0012)	-6.85	< .001***
3	.0014	.0064 (.0014)	-3.76	< .001***
4	.0012	.0067 (.0011)	-5.02	< .001***
5	.0048	.0059 (.0013)	-0.85	.395
6	.0040	.0075 (.0013)	-2.80	.005**
7	.0088	.0060 (.0012)	2.40	.016*
8	.0023	.0072 (.0014)	-3.57	< .001***
9	.0063	.0068 (.0014)	-0.37	.711

Figures



Figure 1. Still images of the two unfamiliar objects.

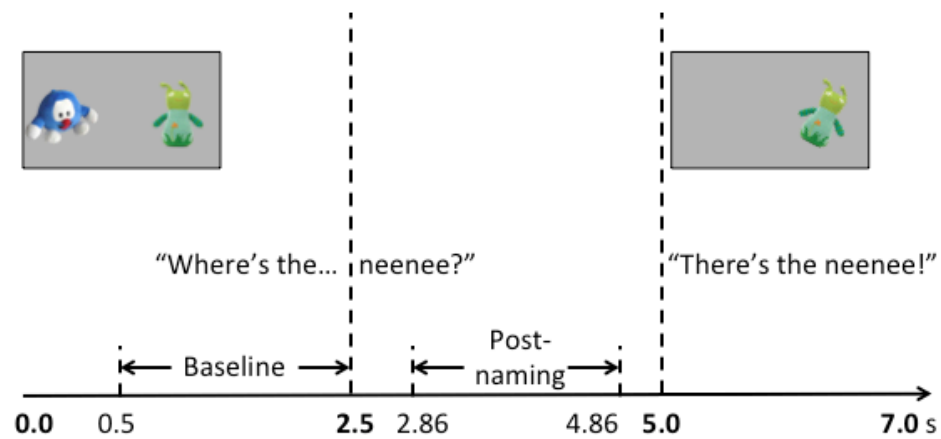


Figure 2. Timeline of a test trial. The test word began 2.5 s after the onset of the trial. At 5.0 s, the distractor image disappeared and the target object started to move. ‘Baseline’ and ‘Post-naming’ refer to the two 2 s windows used in the analysis.

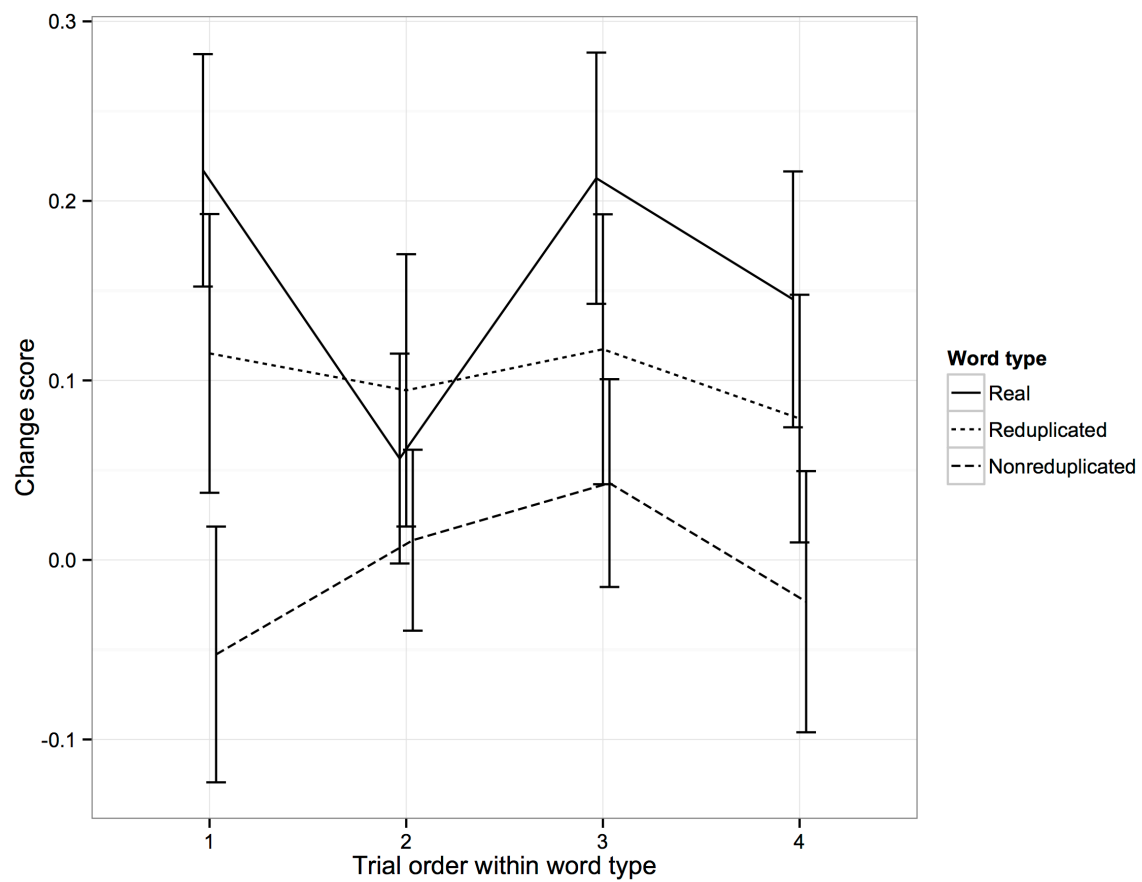


Figure 3. Mean change scores by trial order within word type. Error bars indicate standard errors of mean.

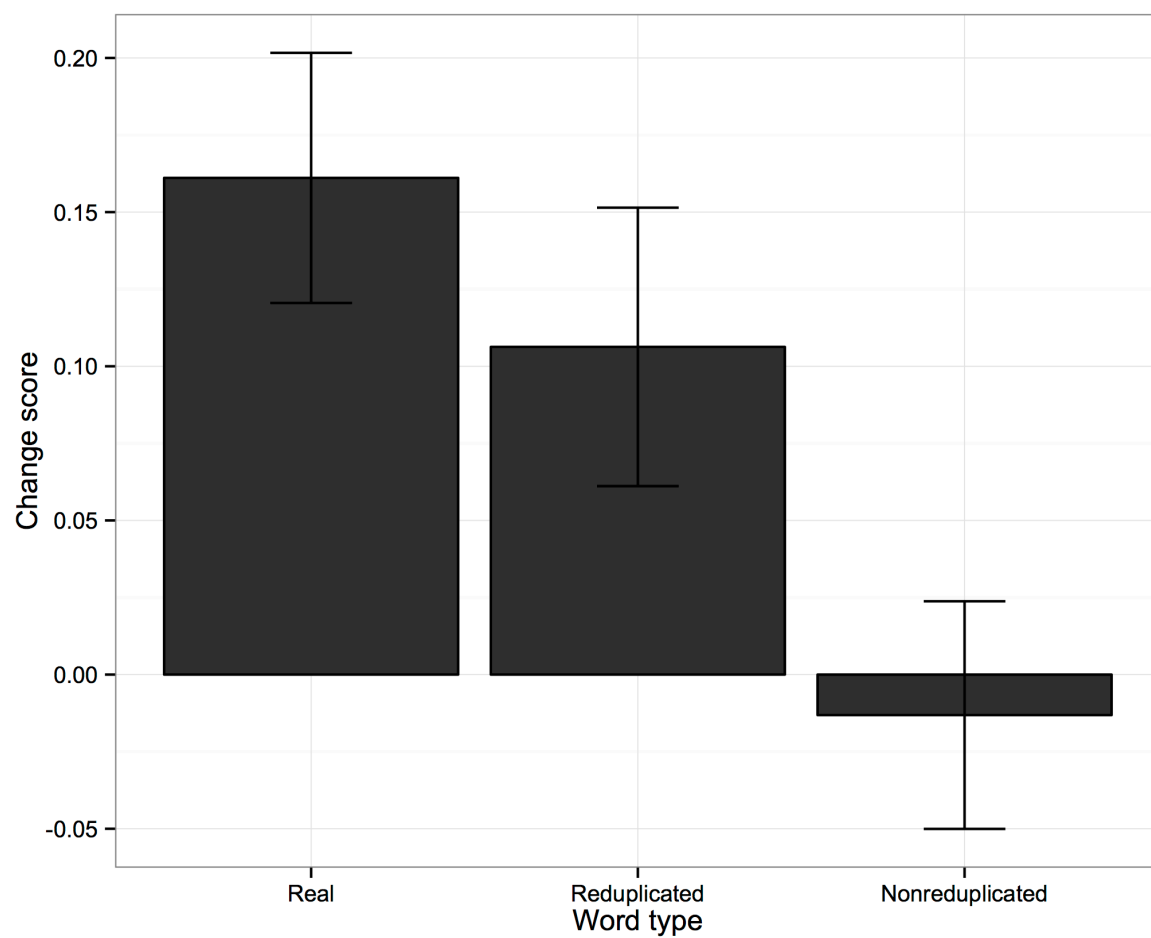


Figure 4. Overall mean change scores by word types. Error bars indicate standard errors of mean.

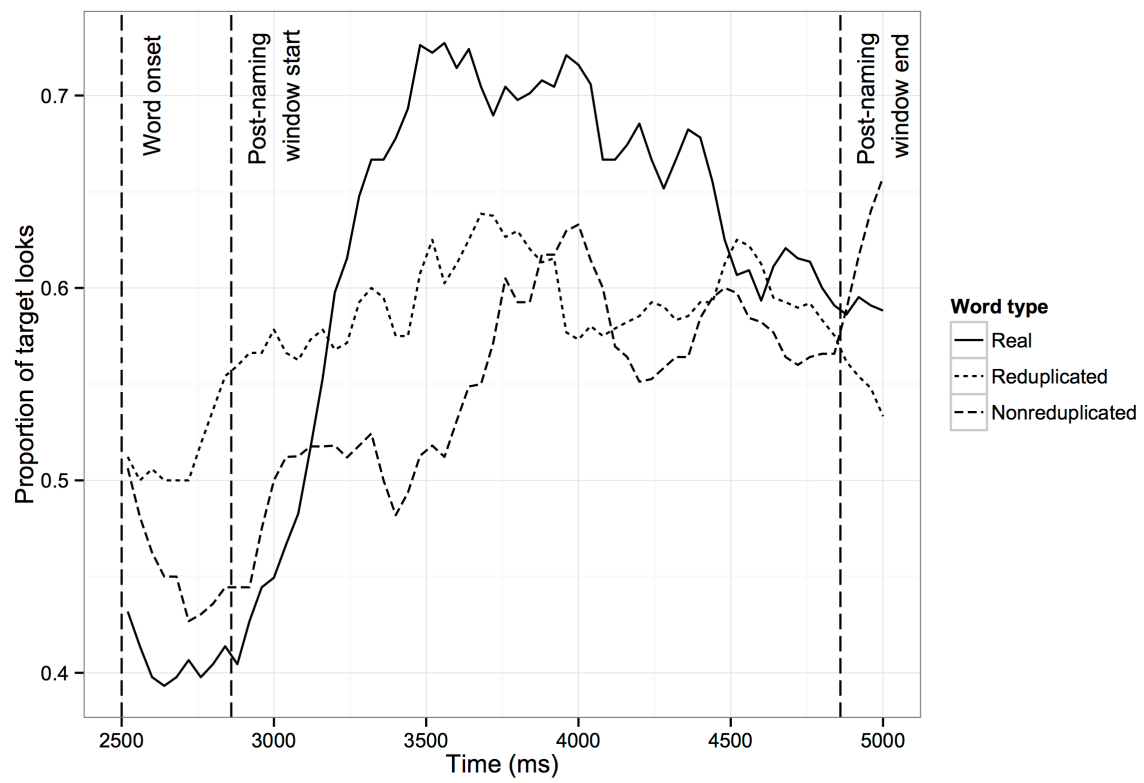


Figure 5. Onset-contingent timecourse plot of target look.

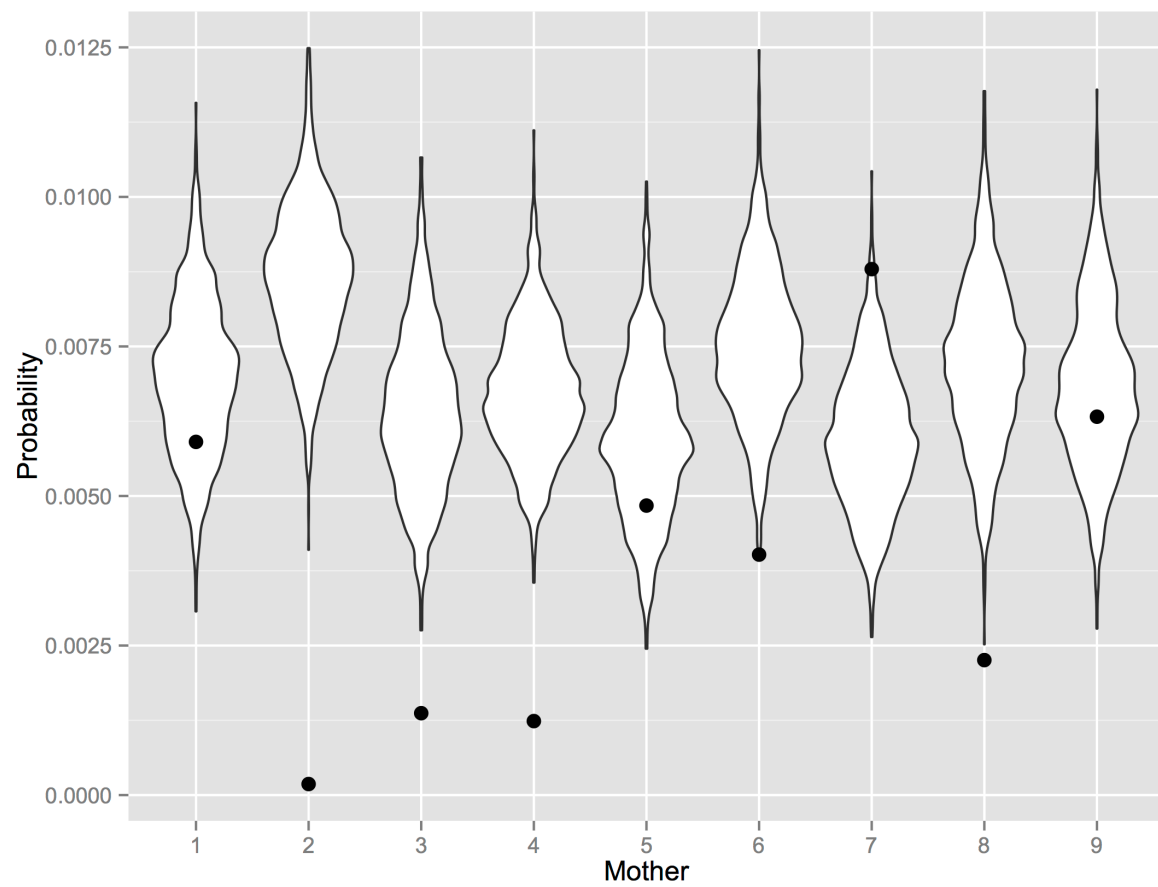


Figure 6. Observed frequencies of syllable reduplication (black dots) plotted against density estimates of simulation results.